Note on some misinterpretations of Gödel's incompleteness theorems

Yong Cheng

Wuhan University, China

16th International Congress on Logic, Methodology And Philosophy of Science and Technology

Prague, 5-10 August 2019

Modern logic and Philosophy

For me, Modern Logic updates and deepens our understanding of the following core concepts of philosophy:

Modern logic and Philosophy

For me, Modern Logic updates and deepens our understanding of the following core concepts of philosophy:

- Truth, Proof, Infinity, Computation, Set, Number, Definability

Modern logic and Philosophy

For me, Modern Logic updates and deepens our understanding of the following core concepts of philosophy:

- Truth, Proof, Infinity, Computation, Set, Number, Definability
- Self-reference, *Incompleteness*, Independence, Decidability

Modern logic and Philosophy

For me, Modern Logic updates and deepens our understanding of the following core concepts of philosophy:

- Truth, Proof, Infinity, Computation, Set, Number, Definability
- Self-reference, *Incompleteness*, Independence, Decidability
- Implication, Consistency, Paradox, Contradiction

Modern logic and Philosophy

For me, Modern Logic updates and deepens our understanding of the following core concepts of philosophy:

- Truth, Proof, Infinity, Computation, Set, Number, Definability
- Self-reference, *Incompleteness*, Independence, Decidability
- Implication, Consistency, Paradox, Contradiction
- Absoluteness, Knowability, Necessity, Vagueness, etc.

Hilbert's program

Two goals of Hilbert's program:

Completeness A proof that all true mathematical statements about arithmetic can be proved in **PA**.

Consistency A proof that no contradiction can be obtained in **PA** using only "finitistic" reasoning about finite mathematical objects.

In this talk, let T be a recursively axiomatized extension of **PA**.

Gödel's incompleteness theorems

Theorem 1 (Gödel, Rosser)

(1) *Gödel's first incompleteness theorem* (G1)*: If T is ω-consistent, then T is not complete (there is a sentence θ such that $T \nvdash \theta$ and $T \nvdash \neg\theta$).*

(2) *Rosser's first incompleteness theorem: If T is consistent, then T is not complete.*

(3) *Gödel's second incompleteness theorem* (G2)*: If T is consistent, then the consistency of T is not provable in T.*

Arithmetization

- The three main ideas in Gödel's proof of G1 are arithmetization, representability and self-reference construction.
- Under arithmetization, any formula or finite sequence of formulas of the theory can be coded by a natural number (called Gödel's number). Under Gödel's arithmetization, the set of Gödel's number of axioms of T is recursive. We use $\ulcorner\phi\urcorner$ to denote the numeral in $L(T)$ of the Gödel number of ϕ.
- Then we could define some relations on natural numbers which express metamathematical property of T. Define $\mathbf{Prf}_T(m, n)$ iff n is the Gödel's number of a proof of the formula with Gödel number m in T. We can show that $\mathbf{Prf}_T(m, n)$ is recursive.

Representability

- A n-ary relation $R(x_1, \cdots, x_n)$ on $\mathbb{N}^n$ is representable in T iff there is a formula $\phi(x_1, \cdots, x_n)$ such that $T \vdash \phi(\overline{m_1}, \cdots, \overline{m_n})$ if $R(m_1, \cdots, m_n)$ holds; and $T \vdash \neg\phi(\overline{m_1}, \cdots, \overline{m_n})$ if $R(m_1, \cdots, m_n)$ does not hold.
- Gödel proves that every recursive relation is representable in **PA**. Let $\mathbf{Proof}_T(x, y)$ be the formula which represents $\mathbf{Prf}_T(m, n)$ in **PA**.
- From the representation formula $\mathbf{Proof}_T(x, y)$, we could define the provability predicate $\mathbf{Prov}_T(x)$ as $\mathbf{Prov}_T(x) \triangleq \exists y \mathbf{Proof}_T(x, y)$.
- $\mathbf{Prov}_T(x)$ satisfies the following conditions:
  (1) If $T \vdash \varphi$, then $T \vdash \mathbf{Prov}_T(\ulcorner\varphi\urcorner)$;
  (2) $T \vdash \mathbf{Prov}_T(\ulcorner\varphi\urcorner) \to (\mathbf{Prov}_T(\ulcorner\varphi \to \psi\urcorner) \to \mathbf{Prov}_T(\ulcorner\psi\urcorner))$;
  (3) $T \vdash \mathbf{Prov}_T(\ulcorner\varphi\urcorner) \to \mathbf{Prov}_T(\ulcorner\mathbf{Prov}_T(\ulcorner\varphi\urcorner)\urcorner)$.

Self-reference construction

- Gödel constructs a Gödel sentence $\mathbf{G}$ which asserts its own unprovability in T, i.e. $T \vdash \mathbf{G} \leftrightarrow \neg \mathbf{Prov}_T(\ulcorner \mathbf{G} \urcorner)$.
- Gödel shows that if T is consistent, then $T \nvdash \mathbf{G}$; and if T is ω-consistent, then $T \nvdash \neg \mathbf{G}$.
- Define $\mathbf{Con}(T) \triangleq \neg \mathbf{Prov}_T(\ulcorner 0 \neq 0 \urcorner)$.
- From the above conditions (1)-(3), we can show that $T \vdash \mathbf{Con}(T) \leftrightarrow \mathbf{G}$.
- Thus, G2 holds: if T is consistent, then $T \nvdash \mathbf{Con}(T)$.

Misinterpretation Any theory of arithmetic is incomplete.

Misinterpretation Any theory of arithmetic is incomplete.

- Whether a theory of arithmetic is complete depends on the language of the base theory.
- **PA** is incomplete in the language of $L(\mathbf{0}, \mathbf{S}, +, \cdot)$. But there are respectively recursively axiomatized complete arithmetic theories in the language of $L(\mathbf{0}, \mathbf{S})$, $L(\mathbf{0}, \mathbf{S}, <)$ and $L(\mathbf{0}, \mathbf{S}, <, +)$.
- Containing the arithmetic of multiplication is essential for the proof of G1. For example, Presburger arithmetic is the theory of arithmetic of addition in $L(\mathbf{0}, \mathbf{S}, +)$; but Presburger arithmetic is complete.
- Containing the arithmetic of multiplication is not a sufficient condition for a theory to be incomplete. For example, there is a complete recursively axiomatized theory in the language $L(\mathbf{0}, \mathbf{S}, \cdot)$.

Misinterpretation Since $Th(\mathbb{N}, +, \cdot)$ is undecidable and not recursive axiomatizable, and $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$, we have: $Th(\mathbb{Z}, +, \cdot)$, $Th(\mathbb{Q}, +, \cdot)$ and $Th(\mathbb{R}, +, \cdot)$ are all undecidable and not recursive axiomatizable.

Misinterpretation Since $Th(\mathbb{N}, +, \cdot)$ is undecidable and not recursive axiomatizable, and $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$, we have: $Th(\mathbb{Z}, +, \cdot)$, $Th(\mathbb{Q}, +, \cdot)$ and $Th(\mathbb{R}, +, \cdot)$ are all undecidable and not recursive axiomatizable.

- Since $Th(\mathbb{N}, +, \cdot)$ is interpretable in $Th(\mathbb{Z}, +, \cdot)$, $Th(\mathbb{Z}, +, \cdot)$ is undecidable and hence not recursive axiomatizable.
- Since $Th(\mathbb{N}, +, \cdot)$ is interpretable in $Th(\mathbb{Q}, +, \cdot)$, $Th(\mathbb{Q}, +, \cdot)$ is undecidable and hence not recursive axiomatizable.
- But $Th(\mathbb{R}, +, \cdot)$ is decidable, recursively axiomatizable theory (even if not finitely axiomatizable) and $Th(\mathbb{R}, +, \cdot) = \mathbf{RCF}$ (the theory of real closed field).
- This fact does not contradict G1 since none of $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ is definable in the structure $(\mathbb{R}, +, \cdot)$.

Misinterpretation Any consistent extension of **PA** is incomplete.

Definition 1

Let T be a theory and Γ be a class of formulas.

(1) *T is Σ_n-definable iff there is a Σ_n formula $\phi(x)$ such that n is the Gödel number of a sentence in T if and only if $\mathfrak{N} \models \phi(\overline{n})$ where $\mathfrak{N} = (\mathbb{N}, +, \cdot)$ is the standard model of arithmetic.*

(2) *T is Σ_n-sound if and only if for all Σ_n sentences ϕ, if $T \vdash \phi$, then $\mathfrak{N} \models \phi$.*

(3) *T is Γ-decisive if and only if for all Γ sentences ϕ, either $T \vdash \phi$ or $T \vdash \neg\phi$ holds.*

From G1, if T is a consistent Σ_1-definable extension of **PA**, then T is not Π_1-decisive.

G1 can be extended to arithmetically definable theories

G1 can be generalized to arithmetically definable theories.

Theorem 2 (Kikuchi and Kurahashi)

If T is Σ_{n+1}-definable and Σ_n-sound extension of **PA**, *then T is not Π_{n+1}-decisive.*

G1 can be extended to arithmetically definable theories

G1 can be generalized to arithmetically definable theories.

Theorem 2 (Kikuchi and Kurahashi)

If T is Σ_{n+1}-definable and Σ_n-sound extension of **PA**, *then T is not Π_{n+1}-decisive.*

The optimality of this generalization is shown by Salehi and Seraji.

Theorem 3 (Salehi and Seraji)

There exists a Σ_{n+1}-definable, Σ_{n-1}-sound ($n \geq 1$) and complete theory which extends **PA**.

Thus, it is not true that any consistent extension of **PA** is incomplete.

Different proofs of incompleteness

A proof of G1 is constructive if it explicitly constructs the independent sentence from the base theory by algorithmic ways.

A proof of G1 has the Rosser property if the proof only assumes that T is consistent.

We could classify proofs of Gödel's incompleteness theorems based on the following criterions:

- Proof via proof theoretic method;
- Proof via recursion theoretic method;
- Proof via model theoretic method;
- Proof via arithmetization;
- Proof via the fixed point lemma;
- Proof via logical paradox;
- Proof via constructive method;
- Proof with the Rosser property;
- Independent sentences with real mathematical content.

Properties of Gödel's proof

Gödel's proof of G1 has the following properties:

- uses proof-theoretic method with arithmetization;
- does not use the fixed point lemma;
- formalize the liar paradox;
- the proof is constructive;
- the proof does not have the Rosser property;
- Gödel's sentence is constructed via metamathematical method and has no real mathematical content.

All these characteristics of Gödel's proof of G1 are not necessary conditions for proving G1.

Grzegorczyk proposed the theory of concatenation (**TC**) with no reference to natural numbers and proved that **TC** is essentially incomplete without arithmetization.

Incompleteness is closely related to paradox. Gödel claimed: "Any epistemological antinomy could be used for a similar proof of the existence of undecidable propositions".

Different proofs of incompleteness theorems via paradox:

- **Gödel** Liar Paradox
- **Boolos, Chaitin, Kikuchi, Vopenka, Kurahashi, Sakai, Tanaka** Berry's paradox
- **Kurahashi, Kikuchi, Priest, Cieśliński and Urbaniak** Yablo's Paradox
- **Kritchman-Raz** Unexpected Examination Paradox
- **Cieśliński** Grelling-Nelson's Paradox

Concrete incompleteness

Misinterpretation Gödel's theorem has no relevance to mathematics.

Gödel's proof uses meta-mathematical method and Gödel's sentence has no real mathematical content.

A natural question is then: can we find true sentences not provable in **PA** with real mathematical content?

Harvey Friedman proposed a research program on concrete incompleteness:

the long range impact and significance of ongoing investigations in the foundations of mathematics is going to depend greatly on the extent to which the Incompleteness Phenomena touche normal concrete mathematics.

Examples of mathematically natural independent statements

Question: Can we find an independent sentence of arithmetic with real mathematical contents?

The following are examples of mathematically natural independent statements of **PA**.

Paris-Harrington: Paris-Harrington principle

Kirby and Paris: The Goodstein sequence, The Hercules-Hydra game

Kanamori-McAloon: The Kanamori-McAloon principle

Beklemishev: The Worm principle

Kirby: The flipping principle

Mills: The arboreal statement

Pudlák: P.Pudlák's Principle

Clote: The kiralic and regal principles

Properties of mathematical examples

- All these naturally combinatorial independent principles with real mathematical contents are in fact provably equivalent in **PA** to a certain metamathematical sentence.
- Let $Rfn_{\Sigma_1}(\mathbf{PA})$ denote the sentence which expresses the reflection principle for Σ_1 sentences. People have showed that $\mathbf{PA} \vdash \varphi \leftrightarrow Rfn_{\Sigma_1}(\mathbf{PA})$, where φ is any one of the above independent principles.
- These independent principles are more complex than Gödel's sentence: Gödel's sentence is equivalent to $\mathbf{Con}(\mathbf{PA})$ in **PA**; but all these principles are not only independent of **PA** but also independent of $\mathbf{PA} + \mathbf{Con}(\mathbf{PA})$.

Concrete incompleteness for higher-order arithmetic

Question 1

Can we find a mathematical theorem expressible in Second Order Arithmetic but not provable in Second Order Arithmetic?

Theorem 4 ([2] [1])

There is a concrete mathematical theorem which is expressible in second-order arithmetic, not provable in second-order and third-order arithmetic but provable in fourth-order arithmetic.

Reference:

Yong Cheng. Incompleteness for higher-order arithmetic: An example based on Harrington's Principle. Springer series: Springerbrief in Mathematics, 2019.

The notion of interpretation

- An interpretation of a theory T in a theory S is a mapping from formulas of T to formulas of S that maps all axioms of T to sentences provable in S. If T is interpretable in S, then all sentences provable (refutable) in T are mapped, by the interpretation function, to sentences provable (refutable) in S.
- Interpretability can be accepted as a measure of strength of different theories. If S is interpretable in T but T is not interpretable in S, then we say that S is weaker than T w.r.t. interpretation.

Definition 2

G1 *holds for* T *iff for any recursively axiomatizable consistent theory* S*, if* T *is interpretable in* S*, then* S *is incomplete.*

T is essentially incomplete iff any recursively axiomatizable consistent extension of T is incomplete.

Proposition 1

Let T *be a recursively axiomatizable consistent theory. The followings are equivalent:*

(1) G1 *holds for* T.

(2) T *is essentially incomplete.*

Robinson arithmetic $\mathbf{Q}$

Definition 3 (Tarski, Mostowski and Robinson)

Robinson arithmetic $\mathbf{Q}$ *is defined in the language* $\{\mathbf{0}, \mathbf{S}, +, \cdot\}$ *with the following axioms:*

- Q1 $\forall x \forall y(\mathbf{S}x = \mathbf{S}y \rightarrow x = y)$;
- Q2 $\forall x(\mathbf{S}x \neq \mathbf{0})$;
- Q3 $\forall x(x \neq \mathbf{0} \rightarrow \exists y(x = \mathbf{S}y))$;
- Q4 $\forall x \forall y(x + \mathbf{0} = x)$;
- Q5 $\forall x \forall y(x + \mathbf{S}y = \mathbf{S}(x + y))$;
- Q6 $\forall x(x \cdot \mathbf{0} = \mathbf{0})$;
- Q7 $\forall x \forall y(x \cdot \mathbf{S}y = x \cdot y + x)$.

The system $\mathbf{R}$

Definition 4 (Tarski, Mostowski and Robinson)

The system **R**

Definition 4 (Tarski, Mostowski and Robinson)

(1) *Let* **R** *be the system consisting of schemes* **Ax**1-**Ax**5 *with* $L(\mathbf{R}) = \{\mathbf{0}, \mathbf{S}, +, \cdot, \leq\}$ *where* $m, n \in \mathbb{N}$.

*Ax*1 $\overline{m} + \overline{n} = \overline{m+n}$;

*Ax*2 $\overline{m} \neq \overline{n}$ *if* $m \neq n$;

*Ax*3 $\overline{m} \cdot \overline{n} = \overline{m \cdot n}$;

*Ax*4 $\forall x(x \leq \overline{n} \to x = \overline{0} \vee \cdots \vee x = \overline{n})$;

*Ax*5 $\forall x(x \leq \overline{n} \vee \overline{n} \leq x)$.

The limit of applicability of G1

- G1 holds for **Q**;
- G1 holds for **R**;
- **R** is weaker than **Q** w.r.t. interpretation.

Misinterpretation **R** is the weakest theory w.r.t. interpretation such that G1 holds for **R**.

Recently, I show that there are many theories T such that G1 holds for T and T is weaker than **R** w.r.t. interpretation.

Theorem 5

For any recursively inseparable pair $\langle A, B\rangle$, there is a theory $U_{\langle A,B\rangle}$ such that G1 *holds for $U_{\langle A,B\rangle}$ and $U_{\langle A,B\rangle}$ is weaker than* **R** *w.r.t. interpretation.*

Three comments on G2

- $\mathbf{Con}(T)$ is not provable in T for any recursively axiomatized consistent extension T of $\mathbf{PA}$, but T is reflective: $\mathbf{Con}(S)$ is provable in T for any finite sub-theory S of T.
- Only assuming that T is consistent, we cannot get that $\mathbf{Con}(T)$ is independent of T: it is not enough to show that $T \nvdash \neg\mathbf{Con}(T)$ only assuming T is consistent. But we could prove that $\mathbf{Con}(T)$ is independent of T by assuming that T is 1-consistent.
- The notion of consistency is not absolute. A theory may be consistent from the external perspective but inconsistent from the internal perspective. For example, let $T = \mathbf{PA} + \neg\mathbf{Con}(\mathbf{PA})$. From G2, T is consistent from the external perspective. But since $T \vdash \neg\mathbf{Con}(T)$, T is not consistent from the internal perspective of T.

The intensionality of G2

- We say that G2 holds for T if the consistency statement of T is not provable in T.
- This definition is vague: what do we mean "the consistency statement of T is not provable in T"?
- G2 is essentially different from G1 due to the intensionality of G2: whether G2 holds for T depends on how we formulate the consistency statement.

Factors affecting G2

"Whether G2 holds for T" depends on the following factors:

(1) the definition of provability predicate;

(2) the choice of an arithmetic formula to express consistency;

(3) the choice of the base proof system;

(4) the choice of numberings;

(5) the choice of a specific formula numerating (or representing) the axiom set.

Canonical provability predicate and canonical consistency statement

We say that a formula $\mathbf{Prf}_T(x,y)$ is a proof predicate of T if $\mathbf{Prf}_T(x,y)$ satisfies the following conditions:

(1) $\mathbf{Prf}_T(x,y)$ is primitive recursive;

(2) $\mathbf{PA} \vdash \forall x(\mathbf{Prov}_T(x) \leftrightarrow \exists y \mathbf{Prf}_T(x,y))$;

(3) for any $n \in \omega$ and formula ϕ, $N \models \mathbf{Proof}_T(\ulcorner \phi \urcorner, \overline{n}) \leftrightarrow \mathbf{Prf}_T(\ulcorner \phi \urcorner, \overline{n})$;

(4) $\mathbf{PA} \vdash \forall x \forall x' \forall y(\mathbf{Prf}_T(x,y) \wedge \mathbf{Prf}_T(x',y) \to x = x')$.

- Define the canonical provability predicate $\mathbf{Pr}_T(x)$ as $\mathbf{Pr}_T(x) \triangleq \exists y\, \mathbf{Prf}_T(x,y)$.
- Define the canonical consistency statement $\mathbf{Con}(T)$ as $\neg \mathbf{Pr}_T(\ulcorner 0 \neq 0 \urcorner)$.

G2 and the Hilbert-Bernays-Löb Derivability Condition

Definition 5 (Hilbert-Bernays-Löb Derivability Condition)

- D1 *If* $T \vdash \varphi$, *then* $T \vdash \mathbf{Pr}_T(\ulcorner \varphi \urcorner)$;
- D2 $T \vdash \mathbf{Pr}_T(\ulcorner \varphi \urcorner) \rightarrow (\mathbf{Pr}_T(\ulcorner \varphi \rightarrow \psi \urcorner) \rightarrow \mathbf{Pr}_T(\ulcorner \psi \urcorner))$;
- D3 $T \vdash \mathbf{Pr}_T(\ulcorner \varphi \urcorner) \rightarrow \mathbf{Pr}_T(\ulcorner \mathbf{Pr}_T(\ulcorner \varphi \urcorner) \urcorner)$.

Theorem 6 (Löb)

If provability predicate $\mathbf{Pr}_T(x)$ *satisfies* **D1**-**D3**, *then* $T \nvdash \mathbf{Con}(T)$ *if* T *is consistent.*

G2 and the definition of provability predicate

- "Whether G2 holds for T" depends on the choice of provability predicate.
- We say $\mathbf{Pr}_T(x)$ is a standard provability predicate if it satisfies the Hilbert-Bernays-Löb Derivability Condition **D1-D3**.
- G2 holds for standard provability predicates.
- If $\mathbf{Pr}_T(x)$ is not a standard provability predicate, then G2 may fail for T.
- Define the Rosser provability predicate $\mathbf{Pr}^R_T(x)$ as the formula $\exists y(\mathbf{Prf}_T(x,y) \wedge \forall z \leq y \neg \mathbf{Prf}_T(\dot\neg(x), z))$, where $\dot\neg$ is a function symbol expressing a primitive recursive function calculating the code of $\neg\phi$ from the code of ϕ.
- But, G2 always fails for Rosser provability predicate: $T \vdash \mathbf{Con}_R(T) \triangleq \neg\mathbf{Pr}^R_T(\ulcorner 0 \neq 0 \urcorner)$.

G2 and the Hilbert-Bernays derivability condition

Define $\mathbf{Con}^*(T) \triangleq \forall x(\mathbf{Fml}(x) \wedge \mathbf{Pr}_T(x) \rightarrow \neg \mathbf{Pr}_T(\dot{\neg} x))$.

Definition 6

HB1: *If* $T \vdash \phi \rightarrow \varphi$, *then* $T \vdash \mathbf{Pr}_T(\ulcorner \phi \urcorner) \rightarrow \mathbf{Pr}_T(\ulcorner \varphi \urcorner)$.

HB2: $T \vdash \mathbf{Pr}_T(\ulcorner \neg \phi(x) \urcorner) \rightarrow \mathbf{Pr}_T(\ulcorner \neg \phi(\dot{x}) \urcorner)$.

HB3: $T \vdash f(x) = 0 \rightarrow \mathbf{Pr}_T(\ulcorner f(\dot{x}) = 0 \urcorner)$ *for every primitive recursive term* $f(x)$.

HB1-HB3 is called the Hilbert-Bernays derivability condition.

Theorem 7 (Hilbert-Bernays)

If provability predicate $\mathbf{Pr}_T(x)$ *satisfies* **HB1-HB3**, *then* $T \nvdash \mathbf{Con}^*(T)$ *if* T *is consistent.*

G2 and the choice of arithmetic formula to express consistency

Theorem 8 (Kurahashi)

There exists a Rosser provability predicate satisfying the Hilbert-Bernays derivability condition.

Thus, the canonical consistency statement based on this Rosser provability predicate is provable: $T \vdash \mathbf{Con}_R(T)$.

But the consistency statement based on this Rosser provability predicate is not provable: $T \nvdash \mathbf{Con}^*_R(T)$.

G2 and the choice of base system

An foundational question about G2 is: how much of information about arithmetic is required for the proof of G2. If the base proof system does not contain enough information about arithmetic, then G2 may fail.

Bezboruah and Shepherdson define the consistency of **Q** as the sentence $\mathbf{Con}^0(\mathbf{Q})$ and show that $\mathbf{Q} \nvdash \mathbf{Con}^0(\mathbf{Q})$.

A natural question is whether G2 holds for theories weaker than **Q** w.r.t. interpretation (e.g. whether G2 holds for **R**).

Dan Willard has constructed examples of r.e. arithmetical theories that couldn't prove the totality of successor function but could prove their own canonical consistency statement.

Fedor Pakhomov defined a theory $H_{<\omega}$ and showed that it proves its own canonical consistency statement.

G2 and the choice of numberings

Any injective function γ from a set of expressions in $L(\mathbf{PA})$ to $\mathbb{N}$ qualifies as a numbering.

Gödel's numbering is a special kind of numberings under which the Gödel number of the set of axioms of **PA** is recursive.

"Whether G2 holds for T" depends on the choice of numbering.

Grabmayr examines different criterions of acceptability and shows that G2 holds for acceptable numberings.

If the numbering is not acceptable, then G2 may fail. Grabmayr gives some examples of deviant numberings γ such that G2 fails.

Provability predicate via numeration

Now we define provability predicate and consistency statement w.r.t. numerations.

Definition 7

Let $\alpha(x)$ be a formula in $L(T)$.

(1) *$\alpha(x)$ is a numeration of T if for any n, $\mathbf{PA} \vdash \alpha(\overline{n})$ iff n is the Gödel number of some sentence in T.*

(2) *Define the formula $\mathbf{Prf}_{\alpha}(x, y)$ saying "y is the Gödel number of a proof of the formula with Gödel number x from the set of all sentences satisfying $\alpha(x)$".*

(3) *Define the provability predicate $\mathbf{Pr}_{\alpha}(x)$ of $\alpha(x)$ as $\mathbf{Pr}_{\alpha}(x) \triangleq \exists y \mathbf{Prf}_{\alpha}(x, y)$ and consistency statement $\mathbf{Con}_{\alpha}(T)$ as $\triangleq \neg\mathbf{Pr}_{\alpha}(\ulcorner 0 \neq 0 \urcorner)$.*

G2 and the numeration of base theory

Theorem 9 (Generalization of G2)

Let T be any recursively enumerable consistent extension of **PA**. *If $\alpha(x)$ is any Σ_1 numeration of T, then $T \nvdash \mathbf{Con}_{\alpha}(T)$.*

- G2 holds for Σ_1 numerations of **PA**, but fails for some Π_1 numerations of **PA**.
- Feferman showed that if T is a r.e. consistent extension of **PA**, then there is a Π_1 numeration $\tau(u)$ of T such that $T \vdash \mathbf{Con}_{\tau}(T)$.

Gödel's theorems and the mechanism thesis

The mechanism thesis: the mind cannot be mechanized (i.e. the mathematical outputs of the idealized human mind outstrip the mathematical outputs of any Turing machine).

A popular interpretation of G1 is that it implies that the mind cannot be mechanized.

I will not examine the broad question of whether the mind can be mechanized. I will only examine the question of whether G1 implies that the mind cannot be mechanized.

Gödel did not argue that his theorems imply that the mind cannot be mechanized; instead he argued that his theorems imply a weaker conclusion: Gödel's Disjunction.

Gödel's Disjunction

The first disjunct (FD) The mind cannot be mechanized;

The second disjunct (SD) There are absolutely undecidable statements in the sense that there are mathematical truths that cannot be proved by the idealized human mind.

Gödel's Disjunction (GD) Either the first disjunct holds or the second disjunct holds.

Gödel's disjunctive thesis

- Lucas and Penrose argue for the first disjunct based on Gödel's theorems.
- For Gödel, the first disjunct is true (the mind cannot be mechanized) and the second disjunct is false (human mind is sufficiently powerful to capture all mathematical truths).
- For Gödel, **GD** is a mathematically established fact of great philosophical interest following from Gödel's theorems.

Epistemic Arithmetic $\mathbf{EA_T}$

Let $\mathbf{K}$ be the set of sentences in $L(\mathbf{PA})$ that the idealized human mind can know. Let $\langle M_e : e \in \mathbb{N}\rangle$ be an enumeration of Turing machines and $Th(M_e)$ be the theory enumerated by the Turing machine M_e.

People have developed $\mathbf{EA_T}$, a system of epistemic arithmetic, which consists of axioms of arithmetic, axioms of the absolute knowability operator $\mathbf{K}$ and axioms of typed truth.

Theorem 10

(Reinhardt) Assume $\mathbf{EA_T}$. *Then* GD *holds.*

Three levels of the mechanistic thesis

The weak mechanistic thesis (WMT) $\exists e(\mathbf{K} = Th(M_e))$;

The strong mechanistic thesis (SMT) $\mathbf{K}\exists e(\mathbf{K} = Th(M_e))$;

The super strong mechanistic thesis (SSMT)

$$\exists e\,\mathbf{K}(\mathbf{K} = Th(M_e)).$$

- WMT is just the first disjunct which says that there is a Turing machine which coincides with the idealized human mind in the sense that the two have the same outputs.
- SMT says that the idealized human mind knows that there is a Turing machine which coincides with the idealized human mind.
- SSMT says that there is a particular Turing machine such that the idealized human mind knows that that particular machine coincides with the idealized human mind.

Theorem 11

(Reinhardt) $\mathbf{EA_T} + \mathrm{SSMT}$ *is inconsistent.*

We can prove in $\mathbf{EA_T}$ that there does not exist a particular Turing machine such that the idealized human mind knows that that particular Turing machine coincides with the idealized human mind.

Theorem 12

(Reinhardt) $\mathbf{EA_T} + \mathrm{WMT}$ *is consistent.*

From the viewpoint of $\mathbf{EA_T}$, it is possible that the idealized human mind is in fact a Turing machine.

Theorem 13

(Carlson) $\mathbf{EA_T} + \mathrm{SMT}$ *is consistent.*

From the point of view of $\mathbf{EA_T}$, it is possible that the idealized human mind knows that it is a Turing machine: it just cannot know which one.

Penrose's new argument is the most sophisticated and promising argument for the first disjunct which has been extensively discussed and carefully analyzed in the literature.

Koellner and others have found a framework **DTK** which employs Feferman's type-free theory of determinate truth **DT** and some additional axioms governing **K** to the axioms of **DT**.

(1) **DTK** is consistent.

(2) **DTK** proves GD.

(3) The particular argument Penrose gives for the first disjunct fails in **DTK**.

(4) The first disjunct is independent of **DTK**.

(5) The second disjunct is independent of **DTK**.

Reference list

Yong Cheng. Incompleteness for higher-order arithmetic: An example based on Harrington's Principle. Springer series: Springerbrief in Mathematics, 2019.

Yong Cheng and Ralf Schindler. Harrington's Principle in higher order arithmetic. *The Journal of Symbolic Logic,* Volume 80,Issue 02, June 2015, pp 477-489.

Yong Cheng, Note on some misinterpretations of Gödel's incompleteness theorems, preprint.

Yong Cheng. Finding the limit of incompleteness I, preprint.

Feferman, S., Arithmetization of metamathematics in a general setting. Fundamenta Mathematicae 49:35 – 92, 1960.

Friedman, Harvey.: Boolean relation theory and incompleteness. Manuscript, to appear.

Hájek, P. and Pudlák, P., 1993, Metamathematics of First-Order Arithmetic, Berlin: Springer.

Roman Murawski. Recursive Functions and Metamathematics: Problems of Completeness and Decidability, Gödel's Theorems. Springer Netherlands, 1999.

J. R. Lucas. Minds, machines, and Gödel. Philosophy 36 (1961), pp. 120-124.

Roger Penrose. Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford University Press (1994).

Peter Koellner. Gödel's Disjunction. Charpter in Gödel's Disjunction: The scope and limits of mathematical knowledge, edited by Leon Horsten and Philip Welch, Oxford University Press 2016.

Peter Koellner. On the Question of Whether the Mind Can Be Mechanized Part I and II: From Gödel to Penrose. to appear in Journal of Philosophy, 2018.

Thanks for your attention!